Journal of Environmental Economics and Management 64 (2012) 117–135

Contents lists available at SciVerse ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

# On modeling pollution-generating technologies ☆

Sushama Murty [a,*], R. Robert Russell [b], Steven B. Levkoff [b]

[a] Department of Economics, University of Exeter Business School, University of Exeter, United Kingdom
[b] Department of Economics, University of California, Riverside, United States

ARTICLE INFO

*Article history:*
Received 12 March 2011
Available online 10 April 2012

*Keywords:*
Pollution-generating technologies
Free disposability
Weak disposability
Data envelopment analysis
Environmental and technical efficiency measurement

ABSTRACT

We argue analytically that many commonly used models of pollution-generating technologies, which treat pollution as a freely disposable input or as a weakly disposable and null-joint output, may generate unacceptable implications for the trade-offs among inputs, outputs, and pollution. We show that the correct trade-offs in production are best captured if a pollution-generating technology is modeled as an intersection of an intended-production technology of the firm and nature's residual-generation set. The former satisfies standard disposability properties, while the latter violates free (strong) disposability of pollution and pollution-causing inputs. As a result, the intersection—which we call a by-production technology—violates standard free disposability of pollution and pollution-causing inputs. Employing data envelopment analysis on an electric-power-plant database, we illustrate shortcomings, under by-production, of two popular efficiency indexes: the hyperbolic and directional-distance-function indexes. We propose and implement an alternative index with superior properties. Under by-production, most efficiency indexes decompose very naturally into intended-production and environmental efficiency indexes. This decomposition is difficult to find under alternative specifications of pollution-generating technologies.



## 1. Introduction

Our reading of the environmental economics literature reveals three broad features of pollution that economists aim to capture. First, the generation of pollution/residuals seems to proceed hand-in-hand with the processes of consumption and production. Second, the residuals so generated require the use of the assimilative capacity of the environment for their disposal. Third, the generation of the residuals and the consequent use of environmental resources for their disposal generate external effects on both consumers and producers and hence the need for policies to regulate the generation of pollution.

☆ We are indebted to Bob Chambers, Finn Forsund, and M.N. Murty for encouraging us to finish this paper, started years ago. The theory developed in the paper first appeared as a Department of Economics, University of California, Riverside Discussion Paper [28]. The current version includes an empirical application of this theory. The paper has been substantially improved in response to the excellent critiques and suggestions of the Editor, Managing Editor, and referees of this journal. It has also benefitted from discussions with colleagues during presentations at the University of New South Wales, the Efficiency and Productivity Workshop (Auckland University Technology, February 2010), the North American Productivity Workshop (Houston University, June 2010), the University of Exeter (May 2011), and the European Workshop on Efficiency and Productivity (Verona University, July 2011). We are also grateful to Carl Pasurka for providing us with the data used in our empirical application.

* Corresponding author. Fax: + 44 1392 263242.
*E-mail addresses:* s.murty@exeter.ac.uk (S. Murty), robert.russell@ucr.edu (R. Robert Russell), levkoff@gmail.com (S.B. Levkoff).

0095-0696 © 2012 Elsevier Inc. Open access under CC BY-NC-ND license.
http://dx.doi.org/10.1016/j.jeem.2012.02.005

In this paper, we confine ourselves to addressing the first feature alone.[1] In particular, we focus on pollution generated by firms. We distinguish between outputs that firms intend to produce and outputs that are unintentionally (incidentally) generated by firms when they engage in the production of intended outputs. Pollution is such an unintended output. We are mainly concerned with studying the specification of the technology set that best captures the link between production of outputs intended by firms and the generation of pollution.

It is reasonable to say that, in the case of pollution generated by firms, there are some specific aspects about the process of transformation of inputs into intended outputs (*e.g.,* the use of inputs such as coal or the production of outputs such as varieties of cheese that release a strong odor) that trigger additional reactions in nature and (abstracting from abatement activities) *inevitably* result in the generation of pollution as a by-product. In this paper, we refer to these natural reactions, which occur alongside intended production by firms, as *by-production*.

In the case of technologies exhibiting by-production, we observe the inevitability of a certain *minimal* amount of the by-product, given the quantities of certain inputs and/or certain intended outputs. Inefficiencies in production could generate more than this minimal amount of the unintended output. At the same time, we also observe the usual menu of *maximal* possible vectors of intended outputs, given an input vector. Such a menu generally reflects the negative tradeoffs in the production of intended outputs when inputs are held fixed, as production of each of these commodities is costly in terms of the inputs used. Inefficiencies in intended production may imply that less than this maximal amount may get produced. An increase in the amounts of the inputs used increases the menu of intended-output vectors that are technologically feasible. At the same time, it increases the minimal amount of the unintended output that can be generated.[2]

The above underscores two crucial points to note about pollution-generating technologies:

(i) technologies of pollution-generating firms do not satisfy free disposability of by-products (pollution cannot be disposed of below the minimal level described above if inputs and intended outputs are held fixed) and
(ii) in such technologies there is a mutual interdependence between changes in inputs, intended outputs, and pollution—an interdependence that we will argue is more correlation than causation.

In most of the existing literature, the standard building block employed in constructing pollution-generating technologies is the positive correlation between intended and unintended outputs. This literature attributes the observed positive correlation to abatement activities by firms rather than directly to the phenomenon of by-production. Abatement activities of firms involve a diversion of resources (inputs) to mitigate or clean up the pollution they produce. In this paper, we model abatement activities as outputs of the firm. The production of these abatement activities is hence costly, given fixed amounts of resources: the more resources are diverted to abatement activities, the less they are available for producing intended outputs. Hence, an increase in the level of abatement activities leads concomitantly to both lower residual generation and lower production of intended output.

In this literature, however, abatement activities are not usually explicitly modeled as another set of outputs produced by firms.[3] Rather, what is proposed is a "reduced form" of the technology in the space of inputs, by-products, and intended outputs. Special assumptions are made to allow the technology to exhibit a positive correlation between by-products and intended outputs, which is implicitly explained by abatement options open to firms. At the same time, it is also assumed that the technology satisfies the standard disposability assumptions with respect to *all* inputs and intended outputs. The approaches taken in the literature to model the positive correlation include: (a) treating pollution as a standard input, so that the technology satisfies (strong) input free disposability with respect to pollution,[4] or (b) treating pollution as an output but with the technology satisfying the assumptions of weak disposability and null-jointness with respect to intended and unintended outputs.[5] In empirical work, both parametric and non-parametric specifications of such technologies are often employed for measuring technical efficiency, marginal abatement cost, productivity, and growth when economic units also produce incidental outputs like pollution. Both Data Envelopment Analysis (DEA)[6] and econometric approaches are employed in this literature.[7]

We propose a model of pollution-generating technologies that captures the salient features (i) and (ii) of the phenomenon of by-production identified above. Our model of the technology, which we refer to as a "by-production technology," is obtained as a composition of two technologies: an intended-production technology and a residual-generation technology. The former is a standard technology that describes how inputs are transformed into intended

[1] See Murty [26] for a general equilibrium study of the second feature in the light of the first feature.

[2] *E.g.*, a greater amount of usage of coal increases the quantity generated of both smoke and electricity.

[3] For exceptions, see Barbera and McConnell [2], Pethig [30], and Hua and Bien [22].

[4] See, *e.g.,* Baumol and Oates [3], Cropper and Oates [8], and Reinhard et al. [32,33].

[5] Weak disposability is satisfied if unintended outputs can only be contracted in tandem with intended outputs. Null-jointness is satisfied if null pollution implies null output. See Section 4 for formal definitions of these concepts.

[6] DEA employs mathematical programming methods to construct the technology by enveloping the data in the "tightest fitting" convex (sometimes conical) set. See Färe et al. [11] for a basic description of DEA and Fried et al. [19] for surveys of more recent developments.

[7] For measurement issues based on parametric specifications of a technology that treat by-products as outputs and employ weak disposability and null-jointness, see, *e.g.,* Pittman [31], Färe et al. [13], Coggins and Swinton [7], Hailu and Veeman [21], Murty and Kumar [23,24], and Murty et al. [25]. For non-parametric, set-theoretic approaches under similar assumptions on the technology see, *e.g.,* Färe et al. [15], Färe et al. [12], Färe et al. [14], and Boyd and McClelland [4]. (These citations by no means constitute an exhaustive or even a representative list. See Zhou et al. [38] for a comprehensive survey of over a hundred papers employing this approach to the modeling of pollution-generating technologies.)

outputs in production. The latter reflects nature's residual generation mechanism, which is a relationship between pollution (an output) and commodities that cause pollution. Thus, if we assume that it is some inputs (*e.g.,* coal) that cause pollution, then an increase in the use of these inputs results (under standard assumptions) in an increase in intended outputs (say electricity). At the same time, such an increase in the use of these inputs causes also an increase in pollution via nature's residual-generating technology. Thus, even without any reference to explicit abatement efforts by firms, the model generates a positive correlation between pollution generation and intended outputs.

We show that abatement options available to firms can also be explicitly factored into our model. When they are available, they form a part of both the intended-production technology (as their production is also costly in terms of resources/inputs of the firm) and the residual-generation mechanism (as they mitigate pollution). We argue, moreover, that the presence of abatement options implies that data generated by pollution-generating technologies can violate the null-jointness assumption (positive levels of intended output may be consistent with zero levels of pollution). The weak-disposability restriction on pollution-generating technologies does not preclude regions of *negative* correlation between intended and unintended outputs along the frontier. On the other hand, in the by-production technology we formulate, no such regions of negative correlations are observed.

The intended-production technology satisfies standard free-disposability properties with respect to inputs and intended outputs and is assumed to be independent of the level of pollution. As in Murty [26], nature's residual-generating technology treats pollution as an output that satisfies the assumption of "costly disposability" and violates standard disposability properties with respect to goods that result in (affect) pollution generation. As a result, the by-production technology, which is an intersection of the intended-production technology and nature's residual-generating technology, violates standard disposability with respect to goods that cause (or affect) pollution generation and exhibits costly disposability with respect to pollution. In these ways, our proposed by-production approach is different from the standard input and output approaches to modeling pollution-generating technologies.

Though the contribution of this paper is mainly theoretical, to demonstrate the ready applicability of the by-production approach in empirical work, we also conduct numerical and empirical analyses, each employing DEA methods. We first show how our by-production technology can be constructed as the intersection of two DEA technologies, one for intended production and one for residual generation. This construction lends itself to the analysis of many important issues seen in the environmental economics literature. Here, we focus only on one such application: the calculation of efficiency of individual firms. With the help of a simple example based on an artificial dataset, we show that the sets of (weakly) efficient points obtained from the weak-disposability approach usually employed in the DEA literature and the new by-production approach are generally different (the former will be a larger set of points than the latter).

In the context of by-production, the conventional (in)efficiency indexes decompose nicely into an intended-output efficiency index and an environmental efficiency index. We use our example to show that the common indexes employed in this literature, the hyperbolic index and the directional-distance-function index, are seriously flawed when the technology satisfies by-production. In particular, standard indexes tend to overstate efficiency. We then propose an alternative index, a modification of an index proposed by Färe et al. [10], for measurement of efficiency for by-production technologies. This index corrects for the flaws in the hyperbolic and directional-distance-function indexes. These comparisons of different efficiency indexes under the conventional and by-production approaches are confirmed by application to an actual database for 92 coal-fired electric power firms.

## 2. Single-equation representation of pollution-generating technologies

In this section, we show that a single implicit relation between outputs and inputs is not rich enough to capture, simultaneously, all the trade-offs among commodities that are implied by the phenomenon of by-production. In particular, we review the traditional formulations that treat pollution as any other standard (freely disposable) input. Later, in Sections 4–7, we compare the new approach that we adopt in this paper with the conventional approach that treats pollution as a weakly disposable output.

In order to strip the argument to its barest essentials, we consider a very parsimonious model in which two inputs—one pollution generating, the other not—are employed to produce a single intended output and a single unintended output. All of the following calculations can easily be generalized to incorporate multiple outputs and multiple inputs of both kinds; indeed, we carry out this generalization when we introduce our empirical model in Section 4.

### 2.1. The case without abatement output

We first analyze the single-equation representation assuming the absence of an abatement technology, so that input substitution is the only abatement option. Denote the quantities of the two inputs by $x_1$ and $x_2$, where the latter is the quantity of the pollution-generating input. Denote the quantities of intended and unintended outputs, respectively, by $y$ and $z$.

A single-equation formulation of the pollution-generating technology, an extension of the standard functional representation of a multiple-output technology, is as follows:

$$T = \{ \langle x_1, x_2, y, z \rangle \in \mathbf{R}^4_+ \mid f(x_1, x_2, y, z) \le 0 \},$$

where $f$ is differentiable, with derivatives with respect to inputs and intended outputs given, respectively, by[8]

$$f_i(x_1,x_2,y,z) \le 0, \quad i=1,2,$$

$$f_y(x_1,x_2,y,z) \ge 0. \tag{2.1}$$

These constraints are standard differential restrictions to impose free disposability of, respectively, inputs and the intended output[9]:

$$\langle x_1,x_2,y,z \rangle \in T \wedge \overline{x}_1 \ge x_1 \Rightarrow \langle \overline{x}_1,x,y,z \rangle \in T,$$

$$\langle x_1,x_2,y,z \rangle \in T \wedge \overline{x}_2 \ge x_2 \Rightarrow \langle x_1,\overline{x}_2,y,z \rangle \in T,$$

$$\langle x_1,x_2,y,z \rangle \in T \wedge \overline{y} \le y \Rightarrow \langle x_1,x_2,\overline{y},z \rangle \in T. \tag{2.2}$$

A standard stream of research, building on the original Baumol-Oates [3] formulation, models the unintended output as a conventional input, so that the technology satisfies

$$f_z(x_1,x_2,y,z) \le 0 \tag{2.3}$$

and hence

$$\langle x_1,x_2,y,z \rangle \in T \wedge \overline{z} \ge z \Rightarrow \langle x_1,x_2,y,\overline{z} \rangle \in T. \tag{2.4}$$

Quantity vectors satisfying $f(x_1,x_2,y,z)=0$ are points on the frontier of the technology.[10] Those satisfying $f(x_1,x_2,y,z)<0$ are inefficient: more intended output could be produced with given quantities of inputs and pollution; less pollution could be generated with given intended-output and input quantities; and smaller input quantities could be used to produce the given output quantities, given the pollution level.

Suppose $f_z(\hat{x}_1,\hat{x}_2,\hat{y},\hat{z})<0$ for some $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z} \rangle$ satisfying $f(\hat{x}_1,\hat{x}_2,\hat{y},\hat{z})=0$. Then, from the implicit function theorem, $z$ can be expressed as an explicit function of $x_1$, $x_2$, and $y$ in a local neighborhood around $\langle \hat{x}_1,\hat{x}_2,\hat{y} \rangle$, *i.e.*, $z=\zeta(x_1,x_2,y)$, such that

$$\hat{z}=\zeta(\hat{x}_1,\hat{x}_2,\hat{y}) \quad \text{and} \tag{2.5}$$

$$f(x_1,x_2,y,\zeta(x_1,x_2,y))=0. \tag{2.6}$$

The trade-off between the intended output and the unintended output (with inputs held fixed) implied by the implicit function theorem is

$$\frac{\partial \zeta(x_1,x_2,y)}{\partial y} = -\frac{f_y(x_1,x_2,y,z)}{f_z(x_1,x_2,y,z)} \ge 0. \tag{2.7}$$

The trade-off between the input 2 and the unintended output (with the quantities of the intended output and the other inputs held fixed) is

$$\frac{\partial \zeta(x_1,x_2,y)}{\partial x_2} = -\frac{f_2(x_1,x_2,y,z)}{f_z(x_1,x_2,y,z)} \le 0. \tag{2.8}$$

Noting that all these trade-offs are evaluated at points in the technology set that are weakly technically efficient (that is, $f(x_1,x_2,y,z)=0$), the foregoing formulation of a pollution-generating technology seems to be inconsistent with the phenomenon of by-production for the following reasons:

(a) The existence of the function $\zeta$ satisfying (2.7) as a strict inequality implies that there exists a rich menu (a manifold) of (weakly) technically efficient $\langle y,z \rangle$ combinations, with varying levels of $z$, that are possible *holding both input quantities fixed*. In the absence of an abatement technology, this menu is contrary to phenomenon of by-production, since by-production implies that, at fixed levels of inputs (*e.g.*, coal), there is only *one* (weakly) technically efficient (minimal) level of pollution.
(b) Furthermore, the non-positive trade-off between the pollution-generating input (derived by holding the quantity of the intended output fixed), apparent in (2.8), is inconsistent with by-production, as by-production implies that this trade-off should be non-negative.

How should one interpret the trade-offs observed under single-equation modeling of pollution-generating technologies in the absence of an abatement option? As discussed above, these trade-offs are not reflective of the phenomenon of by-production. Rather, the non-negative trade-off observed in (2.7) between the intended output and pollution and the non-positive trade-off observed in (2.8) between input 2 and pollution reflect the treatment of pollution as any other input

[8] Subscripts on $f$ indicate partial differentiation with respect to the indicated variable.

[9] The symbol $\wedge$ stands for "and".

[10] We adopt the following convention in this paper: A point $\langle x_1,x_2,y,z \rangle \in T$ lies on the frontier of $T$ (or is a weakly efficient point of $T$) if there exists no other point $\langle \overline{x}_1,\overline{x}_2,\overline{y},\overline{z} \rangle \in T$ with $\overline{x}_i < x_i$ for $i=1, 2$, $\overline{y} > y$, and $\overline{z} < z$. A point $\langle x_1,x_2,y,z \rangle \in T$ lies on the efficient frontier of $T$ (or is an efficient point of $T$) if there exists no other point $\langle \overline{x},\overline{y},\overline{z} \rangle \in T$ with $\overline{x}_i \le x_i$ for $i=1$, 2, $\overline{y} \ge y$, and $\overline{z} \le z$.

in production: first, increases in its level, holding all other inputs fixed, increases intended output and, second, pollution is a substitute for other inputs in intended production—the same level of intended output can be produced by decreasing other inputs and increasing pollution. This also does not seem to be intuitively correct: it is not a correct description of the role pollution plays in intended production, even if one interprets pollution as the employment of the waste disposal capacity of the environment. This capacity is not usually an input that can substitute for other inputs in production—rather, it is strictly complementary to the other inputs.

### 2.2. The case with abatement output

We model abatement activities as an output, $y^a$, that is used to mitigate pollution.[11] Consider the case where the technology of a pollution-generating firm is defined by a single restriction on all inputs and outputs, including the abatement output[12]:

$$T = \{\langle x_1,x_2,y,z,y^a\rangle \in \mathbf{R}^5_+ \mid f(x_1,x_2,y,z,y^a) \le 0\}. \tag{2.9}$$

We assume that[13]

$$f_a(x_1,x_2,y,z,y^a) \ge 0. \tag{2.10}$$

This restriction captures the fact that the abatement output is also freely disposable,

$$\langle x_1,x_2,y,z,y^a\rangle \in T \wedge \overline{y}^a \le y^a \Rightarrow \langle x_1,x_2,y,z,\overline{y}^a\rangle \in T, \tag{2.11}$$

so that producing it is costly in terms of input usage, implying a non-positive trade-off between it and the other intended outputs. In that case, the implicit function theorem can again be invoked to show that the trade-off between the abatement output and pollution, evaluated in a neighborhood of a (weakly) technically efficient point $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a\rangle \in \mathbf{R}^5_+$ such that $f(\hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a) = 0$ and $f_z(\hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a) < 0$, is

$$\frac{\partial \zeta(x_1,x_2,y,y_a)}{\partial y^a} = -\frac{f_a(x_1,x_2,y,z,y^a)}{f_z(x_1,x_2,y,z,y^a)} \ge 0, \tag{2.12}$$

whenever $f(x_1,x_2,y,z,y^a) = 0$, contradicting the fact that abatement output is produced by firms to mitigate, and not to enhance, pollution.

## 3. A by-production approach to modeling pollution

Given the above analysis, a sound foundation must be identified for introducing multiple production relations to adequately capture the features of by-production. We feel that the resolution to the problem lies in early work of Frisch [20] on production theory, in which he envisaged situations where the correct functional representation of a production technology may require more than one implicit functional relation between inputs and outputs. More recently, Førsund [18] explores these ideas of Frisch. In this section, we build on the works of Frisch and Førsund and show that the phenomenon of by-production requires explicitly distinguishing the by-product-generating mechanism from the intended-production relation. We show that, when this is done, the inconsistencies among trade-offs elucidated in Section 2 get resolved. We also compare our by-production approach with another non-conventional approach (see [29]) based on the material-balance condition in nature.

### 3.1. The by-production approach

In the by-production approach to modeling pollution-generating technologies, the production of the intended output sets a residual-generation mechanism in motion, leading to the generation of the by-product. The analysis can once again be split into two cases: (a) where abatement options are not available to the firm and (b) where the firm has explicit abatement options. Since the analysis is similar in both cases, we focus on the more general case with abatement. Specify the technology as

$$T = T_1 \cap T_2, \text{ where} \tag{3.1}$$

$$T_1 = \{\langle x_1,x_2,y,z,y^a\rangle \in \mathbf{R}^5_+ \mid f(x_1,x_2,y,y^a) \le 0\},$$

$$T_2 = \{\langle x_1,x_2,y,z,y^a\rangle \in \mathbf{R}^5_+ \mid z \ge g(x_2,y^a)\}, \tag{3.2}$$

[11] Examples are end-of-pipe treatment plants (that treat and clean water to remove the pollutant) and production of outputs like scrubbers (which reduce sulfur emissions). As a special case—common in much of the environmental literature—abatement output could be interpreted as abatement itself, in which case *gross* production of the unintended output (before abatement) would be $z+y^a$.

[12] We abstract from long-run abatement options of development, purchase, and installation of new technologies that generate less pollution. See *e.g.*,: Barbera and McConnell [2], where abatement activities include both a purchase of abatement capital and a diversion of some amounts of the usual inputs of a firm towards running of the abatement capital.

[13] $f_a$ denotes the partial derivative of $f$ with respect to the abatement output $y^a$.

and $f$ and $g$ are continuously differentiable functions.[14] The set $T_1$ is a standard technology set, reflecting the ways in which the inputs can be transformed into the intended output and the abatement output. The standard free-disposability properties (2.2) and (2.11) can be imposed on this set by assuming that $f$ satisfies

$$f_i(x_1,x_2,y,y^a) \le 0, \quad i = 1,2,$$

$$f_y(x_1,x_2,y,y^a) \ge 0, \quad \text{and}$$

$$f_a(x_1,x_2,y,y^a) \ge 0. \tag{3.3}$$

Note that $T_1$ imposes no constraint on $z$; that is, it is implicitly assumed that the by-product does not affect the production of intended outputs (formally, $\langle x_1,x_2,y,z\rangle \in T_1 \Rightarrow \langle x_1,x_2,y,\overline{z}\rangle \in T_1 \forall\, \overline{z} \in \mathbf{R}_+$).[15]

The set $T_2$ reflects nature's residual-generation mechanism. We assume that

$$g_2(x_2,y^a) > 0 \quad \text{and} \quad g_a(x_2,y^a) < 0. \tag{3.4}$$

In the formulation of $T_2$, pollution is treated as an output. The restrictions (3.2) and (3.4) capture the fact that pollution is an output of the production process for which disposal is not free. They imply a restriction that is the polar opposite of free output disposability with respect to the unintended output:

$$\langle x_1,x_2,y,z,y^a\rangle \in T_2 \wedge \overline{z} \ge z \wedge \overline{x}_2 \le x_2 \wedge \overline{y}^a \ge y^a \Rightarrow \langle x_1,\overline{x}_2,y,\overline{z},\overline{y}^a\rangle \in T_2. \tag{3.5}$$

Following Murty [26], we refer to this property as "costly disposability" of residuals. Costly disposability implies the possibility of inefficiencies in the generation of pollution (*e.g.*, if a given level of coal generates some minimal level of smoke, then inefficiency in the use of coal may imply that this level of coal can also generate a greater amount of smoke).[16]

The trade-off between $z$ and the pollution-generating input quantity $x_2$ implied by (3.4) is non-negative, and that between $z$ and abatement output $y^a$ is negative. Thus, the sign of $g_a(x_2,y^a)$ captures the mitigating effect abatement has on residual generation and the sign of $g_2(x_2,y^a)$ captures the increase in pollution attributable to the increase in the input causing pollution.

Thus, the overall technology $T$ reflects *both* the transformation of inputs into intended outputs and abatement output (as indicated by the definition of $T_1$ in (3.2) and the use of the abatement output by the firm to control the pollution that results from the use of pollution-generating inputs in intended production (as indicated by the definition of $T_2$ in (3.2). It is easy to infer the disposability properties of $T$ from the above characteristics of $T_1$ and $T_2$:

**Theorem.** *T satisfies free disposability with respect to the intended output and the non-pollution-causing input. It, however, violates free disposability with respect to abatement output and the pollution-causing input. It satisfies costly disposability with respect to pollution.*

The technology violates standard disposability conditions with respect to the quantity of the pollution-causing input $x_2$ because, while $T_1$ satisfies standard free-disposability conditions in $x_2$, $T_2$ satisfies the polar opposite condition with respect to this input. Similarly, $T$ violates standard free disposability of the abatement output because, while $T_1$ satisfies this condition, $T_2$ violates it.

Quantity vectors $\langle x_1,x_2,y,z,y^a\rangle \in T$ that satisfy $f(x_1,x_2,y,y^a) = 0$ and $z = g(x_2,y^a)$ are the weakly efficient points of $T$. If a quantity vector $\langle x_1,x_2,y,z,y^a\rangle \in T$ is such that $f(x_1,x_2,y,y^a) < 0$, then it is technologically possible to decrease the levels of the non-pollution-causing input without changing the production levels of the other input and the output. If a quantity vector in $\langle x_1,x_2,y,z,y^a\rangle \in T$ is such that $z > g(x_2,y^a)$, then it is technologically possible to decrease the level of pollution without changing the production levels of inputs and the intended output.

The implicit function theorem again allows us to derive and sign the local trade-off between pollution and the intended output at any weakly efficient point $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a\rangle$ of $T$. In this simple model, this amounts to first inverting $g$ in $y^a$

$$y^a = g^{-1}(x_2,z) =: h(x_2,z). \tag{3.6}$$

Then substitute into $f(x_1,x_2,y,y^a) = 0$ to obtain

$$\tilde{f}(x_1,x_2,y,z) := f(x_1,x_2,y,h(x_2,z)) = 0. \tag{3.7}$$

[14] The model and the calculations for the case where abatement options are absent are available in Appendix posted on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/.

[15] This could be generalized, of course, allowing pollution to have an effect on intended production as well; *e.g.*, smoke could adversely affect the productivity of labor engaged in producing the intended output. See Murty [27] for a generalization.

[16] There, of course, must be an upper bound as well as a lower bound on pollution for given amounts of inputs and intended outputs. We do not incorporate this upper bound into our model because it is only the lower bound that is of interest for policy makers and researchers, and it is only the lower bound that we construct in the numerical examples and empirical application in Sections 4 and 5 below.

If $f_y(\hat{x}_1,\hat{x}_2,\hat{y},\hat{y}^a)>0$ then, in a local neighborhood around $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z}\rangle$, we can invert (3.7) to obtain the explicit function $y=\psi(x_1,x_2,z)$ and the trade-off between pollution and the intended output at $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a\rangle$ as

$$\frac{\partial\psi(\hat{x}_1,\hat{x}_2,\hat{z})}{\partial z}=-\frac{f_a(\hat{x}_1,\hat{x}_2,\hat{y},\hat{y}^a)h_z(\hat{x}_2,\hat{z})}{f_y(\hat{x}_1,\hat{x}_2,\hat{y},\hat{y}^a)}\geq 0. \tag{3.8}$$

How should one interpret this non-negative "trade-off" between $y$ and $z$? In a local neighborhood of the weakly efficient point $\langle \hat{x}_1,\hat{x}_2,\hat{y},\hat{z},\hat{y}^a\rangle\in T$, an increase in $z$ (holding the levels of both inputs fixed) is attributable (because of the by-production phenomenon inherent in $T_2$) to a reduction in abatement effort $y^a$ by the firm (as $h_z(x_2,z)<0$). Under the conventional assumptions on intended production in (3.3), the trade-off between the abatement-output quantity $y^a$ and intended-output quantity $y$ is

$$-\frac{f_a(x_1,x_2,y,y^a)}{f_y(x_1,x_2,y,y^a)}\leq 0. \tag{3.9}$$

Hence, the reduction in the abatement output implies an increase in resources diverted towards production of other intended outputs $y$. The non-negative trade-off seen in (3.8) between an intended output and pollution at a weakly efficient point of $T$, hence, reflects a non-negative correlation between these commodities effected by the abatement effort of the firm to mitigate by-production of pollution.

### 3.2. A reduced-form representation of the by-production technology

Employing (3.6) and (3.7), we can rewrite the technology in the space of *all* commodities as

$$T=\{\langle x_1,x_2,y,z,y^a\rangle\in\mathbf{R}^5_+\mid \tilde{f}(x_1,x_2,y,z)\leq 0\ \wedge\ y^a\geq h(x_2,z)\}. \tag{3.10}$$

A reduced-form of technology $T$ in the space of the intended output, the unintended output, and inputs can now be derived from (3.10) as

$$\tilde{T}:=\{\langle x_1,x_2,y,z\rangle\in\mathbf{R}^4_+\mid \tilde{f}(x_1,x_2,y,z)\leq 0\}. \tag{3.11}$$

The input and output approaches in the conventional literature model a reduced-form technology—quite in the spirit of $\tilde{T}$—that exhibits a positive correlation between the intended and unintended outputs but satisfies *all* of the standard free disposability assumptions with respect to intended outputs and inputs. The technology is modeled only in reduced form because, although this literature attributes the positive correlation to abatement options available to firms, abatement activities are not explicitly modeled.

As indicated by (3.8), the reduced form of the by-product technology does imply a non-negative trade-off between an intended and an unintended output. However, the derivative of the function $\tilde{f}$ with respect to the pollution-causing input is

$$\tilde{f}_2(x_1,x_2,y,z)=f_a(x_1,x_2,y,y^a)h_2(x_2,z)+f_2(x_1,x_2,y,y^a). \tag{3.12}$$

Given (3.6) and the sign conventions in (3.3) and (3.4), the sign of $\tilde{f}_2$ is ambiguous, contrary to the conventional literature, where it is signed as per a normal input. This corroborates one of the conclusions of our Theorem: that the technology $T$ violates standard free disposability in pollution-causing inputs. To understand this ambiguity, consider the implied trade-off between the intended output and the pollution-causing input

$$-\frac{\tilde{f}_2(x_1,x_2,y,z)}{\tilde{f}_y(x_1,x_2,y,z)}=-\frac{f_a(x_1,x_2,y,y^a)h_2(x_2,z)}{f_y(x_1,x_2,y,y^a)}-\frac{f_2(x_1,x_2,y,y^a)}{f_y(x_1,x_2,y,y^a)}. \tag{3.13}$$

Holding the levels of pollution and the first input fixed, an increase in the level of the pollution-causing input has two opposite effects on intended output: (i) the standard non-negative effect, $-f_2(x_1,x_2,y,y^a)/f_y(x_1,x_2,y,y^a)$, and (ii) a non-positive effect that arises because, to keep the pollution level unchanged when the level of the pollution-causing input is increased, the abatement output must also increase. The extent of this increase is captured by the positive term $h_2(x_2,z)$ in (3.13). But, given fixed resources, the increase in abatement output must come at a cost of a decrease in the level of the intended output ($-f_a(x_1,x_2,y,y^a)/f_y(x_1,x_2,y,y^a)$ is the marginal rate of transformation of the intended output into the abatement output). The term, $-f_a(x_1,x_2,y,y^a)h_2(x_2,z)/f_y(x_1,x_2,y,y^a)$, in (3.13), hence, captures the total reduction in the intended output attributable to the increase in abatement required to hold pollution constant when the quantity of the pollution-generating input is increased.[17]

[17] To give a concrete example: suppose $f(x_1,x_2,y,y^a)=0\Rightarrow y=x_1^{\alpha_1}x_2^{\alpha_2}-y^a$ and $z=g(x_2,y^a)=\beta x_2-\theta y^a$, with $\beta>0$, $\theta>0$, $\alpha_1>0$, and $\alpha_2>0$. Inverting $g$, we obtain $y^a=(\beta x_2-z)/\theta=h(x_2,z)$. Substitution into $f$ yields $y=x_1^{\alpha_1}x_2^{\alpha_2}-(\beta x_2-z)/\theta$. Hence, the trade-off between intended output and the pollution-generating input is $\partial y/\partial x_2=\alpha_2 x_1^{\alpha_1}x_2^{\alpha_2-1}-\beta/\theta x_2$. The sign of this derivative is ambiguous; *e.g.*, it is negative if $x_1=x_2=1$ and $\alpha_2<\beta/\theta$.

### 3.3. The by-production approach and the material-balance condition

Recently, another strand of non-conventional literature, based on the seminal work of Ayres and Kneese [1], has adopted a material-balance approach to modeling pollution-generating technologies.[18] "Material balance" refers to a mass–energy accounting identity derived from the physical law of conservation of mass–energy: the existence of a differential between the mass of material inputs that go into intended production and the mass of intended outputs produced implies that residuals are generated in a production process. Thus, observation of this mass differential can account for residuals of intended production (assuming closedness of the physical system).

Pethig [30] demonstrates that conventional approaches, which attribute the empirically observed positive correlation between pollution generation and intended-output production to diversion of resources by firms to abatement activities *but* fail to model abatement explicitly and consider only a reduced-form technology, violate the material-balance principle.[19] He distinguishes between gross and net residual generation and shows that this problem can be resolved when (i) the gross residuals are explicitly accounted for by the material-balance condition and (ii) abatement activities of the firm, which transform harmful residuals into non-harmful forms, are explicitly modeled along with its intended-production technology. Clearly, as in the by-production model, Pethig's model of a pollution-generating technology is characterized by several production relations.

The by-production approach outlined in Section 3 of this paper demonstrates that the positive correlation between pollution generation and production of intended outputs by firms is a more fundamental phenomenon than is characterized by most of the existing literature, since it is observed even in the absence of explicit abatement options. Residual generation is explained in this approach not as a mass–energy accounting identity, but rather as nature's pollution-generating technology, which operates independently of a firm's intended-production technology. The two technologies are characterized by different sets of input-output relations. But both technologies can have common factors that affect them: *e.g.*, inputs like coal that produce intended outputs like electricity also produce pollution—the output of nature's technology. Thus, a firm engaging in intended production also sets nature's pollution-generating mechanism in action. This is the fundamental explanation of the observed positive correlation. As demonstrated in Sections 3.1 and 3.2, this by-production technology implies that the overall technology of a pollution-generating firm violates standard input free disposability of pollution-causing inputs. Hence, in general, in the corresponding reduced-form technology, the usual sign conventions for derivatives of the production function with respect to inputs *may not* hold in the case of pollution-generating inputs. The particular assumptions made in the material balance approach, however, lead to a special case where the usual sign conventions hold—see Proposition 2 of Pethig [30].

There are two main reasons for the conventional trade-offs of the reduced form in Pethig [30]. First, in Pethig, residuals of the abatement technology are distinguished from residuals of intended production and the sign conventions are derived only with respect to the latter. Second, because residual accounting through a mass-balance condition requires measuring all commodities in common units of mass, the marginal productivities of material inputs in intended production are assumed to be bounded between zero and one—one mass-unit increase in material input results in a less than one mass-unit increase in intended output, the excess being the increase in residual generation. Examples exist where it may not be possible to measure intended outputs and material inputs in common mass units and where "harmless" omissions of some commodities that are not considered relevant by the modeler may result in the violation of the mass-balance condition. The by-production approach, on the other hand, can accommodate approximations and such harmless omissions.[20]

The by-production approach is consistent with the physical laws of conservation of mass and energy provided we make no specification errors in modeling the technology, account for all variables and parameters that describe the rules of residual generation in nature, and make no measurement errors. Specifically, assuming these conditions hold, the relations that define nature's residual-generating mechanism and intended production of a given pollution-generating technology satisfy the physical laws of conservation of mass and energy. Conversely, nature's residual-generating set corresponding to a given pollution-generating technology can be derived from the physical laws of conservation of mass and energy once the relations defining its intended production are specified correctly. Some simple examples that demonstrate these points can be found in Appendix posted on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/. These examples show the extreme case-specificity of modeling pollution-generating technologies. Correct modeling is one that is based on a good engineering understanding of both the relations in nature that define residual generation and the relations that define intended production.

## 4. Data-based pollution-generating technologies

We now demonstrate the ready applicability of the by-production approach to empirical work involving pollution-generating technologies. The current section and Sections 5–7 are devoted to this purpose. Though in this paper we adopt

[18] See, *e.g.*, Pethig [30], Coelli et al. [6], and Chambers and Melkonyan [5].

[19] When a material-balance condition is imposed on these models, they generate a counterintuitive negative trade-off between pollution and intended outputs.

[20] See, for instance, Example A-2 in Appendix posted on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/.

a data envelopment analysis (DEA) approach to constructing pollution-generating technologies, the extension to an econometric approach that models by-production is not difficult to foresee.[21]

We consider the same model as in Section 3, but extend it to incorporate the more general case of multiple inputs, intended outputs, and pollutants. The dataset used below for our empirical application does not contain information on abatement. Hence, in the current section and in Sections 5 and 6 we do not model abatement activity of the firm. In Section 7, however, we consider a numerical example to illustrate the empirical extension to the case with abatement options.

First augment the notation in Sections 2 and 3 as follows:

(i) $p$ decision making units (DMUs),[22] indexed by $d$.
(ii) $m$ intended outputs, indexed by $j$, with quantity vector $y \in \mathbf{R}^m_+$. The $p \times m$ matrix of observations on intended output quantities is denoted by $Y$.
(iii) $n$ inputs, indexed by $i$. The first $n_1$ are non-pollution-generating, while the remaining $n_2 = n - n_1$ are pollution generating. The quantity vector is $x = \langle x^1, x^2 \rangle \in \mathbf{R}^n_+$. The $p \times n$ matrix of observations on the input quantities is denoted by $X = \langle X^1, X^2 \rangle$.
(iv) $m'$ pollutants, indexed by $k$, with quantity vector $z \in \mathbf{R}^{m'}_+$. The $p \times m'$ matrix of observations on pollutants is denoted by $Z$.

For illustrative purposes, we posit an example for a very simple special case with five decision making units, one intended output, one unintended output, and one input:

**Example 1.** $p = 5$, $m = 1$, $n = n_1 = 1$, and $m' = 1$. The (artificial) data are as follows:

| DMU | $x$ | $y$ | $z$ | DMU | $x$ | $y$ | $z$ |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 4 | 4 | 2 | 3 | 5 |
| 2 | 1 | 3/2 | 1 | 5 | 2 | 2 | 3 |
| 3 | 1 | 2/3 | 2 | | | | |

(4.1)

In the conventional output approach to modeling pollution-generating technologies, all intended outputs and inputs are assumed to satisfy standard disposability conditions, but two key assumptions are made regarding the unintended outputs. The first

$$\langle x, y, z \rangle \in \tilde{T} \wedge \lambda \in [0,1] \Rightarrow \langle x, \lambda y, \lambda z \rangle \in \tilde{T} \tag{4.2}$$

is called "weak disposability", a concept originally attributable to Shephard [35,36]. The second

$$\langle x, y, z \rangle \in \tilde{T} \wedge z = 0 \Rightarrow y = 0 \tag{4.3}$$

is called "null-jointness". Weak disposability (WD) and null-jointness imply that (a) while pollution is not freely disposable, it is possible to reduce, in tandem, pollution and the intended outputs and (b) production of *any* positive level of intended output always results in positive amounts of the residual being generated. This literature is predicated on the belief that these two assumptions can capture the fact that, starting at any efficient point of the technology, it is not possible to decrease pollution without decreasing the production of the intended outputs, and hence that, together, they model the positive reduced-form correlation between pollution and other intended outputs. The standard DEA construction of a pollution-generating technology (based on the assumptions of WD and null-jointness), first formulated by Färe and Grosskopf [9] and first empirically implemented by Färe et al. [15], is given by

$$\tilde{T}_{WD} = \{ \langle x, y, z \rangle \in \mathbf{R}^{n+m+m'}_+ \mid \lambda X \le x \wedge \lambda Y \ge y \wedge \lambda Z = z \text{ for some } \lambda \in \mathbf{R}^p_+ \}. \tag{4.4}$$

The production possibility set satisfying WD for Example 1, with $x = 1$, is shown in Panel 4 of Fig. 1. The relevant DMUs are 1, 2, and 3. Frontier points $b$ and $c$ are the $\langle z, y \rangle$ combinations for DMUs 2 and 1, respectively, while vector $a$, corresponding to DMU 3, falls below the frontier.

The by-production approach requires constructing the set $T_1$, which captures the intended-production activities of firms, and the set $T_2$, which captures nature's residual generation. Denote the overall technology $T_1 \cap T_2$ that satisfies by-production by $T_{BP}$. We assume that $T_1$ satisfies free disposability of inputs and intended outputs (as defined in (2.2)), and that it is closed and convex and satisfies constant returns to scale.[23] The intended-output technology $T_1$ that satisfies these assumptions is obtained in a standard way using DEA techniques as follows:

$$T_1 = \{ \langle x, y, z \rangle \in \mathbf{R}^{n+m+m'}_+ \mid \lambda X \le x \wedge \lambda Y \ge y \text{ for some } \lambda \in \mathbf{R}^p_+ \}. \tag{4.5}$$

[21] The econometric approach must involve simultaneous estimation of two (or more) structural production relations that have the above features. In particular the production relation associated with intended production will be the upper frontier of $T_1$ and the production relation associated with residual generation will be the lower frontier of $T_2$. These production relations should satisfy the trade-offs implied by (3.3) and (3.4).

[22] Here we follow the standard nomenclature in the literature on technical efficiency measurement. The generic DMU could be a firm, a plant belonging to a specific firm, or any of a number of types of units of study.

[23] In addition, $T_1$ satisfies independence of $T_1$ from $z$; see (3.2) and footnote 15. The returns-to-scale assumption can be generalized along the lines of Färe et al. [11].

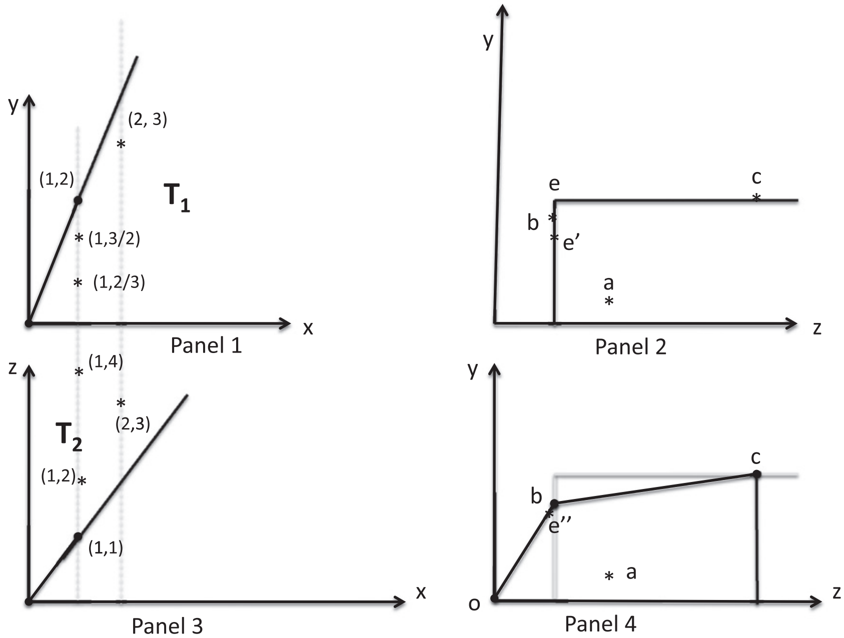


**Fig. 1.**

We assume $T_2$ satisfies costly disposability of pollution and inputs that cause pollution:

$$\langle x^1,x^2,y,z\rangle \in T_2 \wedge \bar{z} \geq z \wedge \bar{x}^2 \leq x^2 \Rightarrow \langle x^1,\bar{x}^2,y,\bar{z}\rangle \in T_2. \tag{4.6}$$

The DEA version of $T_2$, which satisfies these assumptions and exhibits constant returns to scale is obtained as

$$T_2 = \{\langle x^1,x^2,y,z\rangle \in \mathbf{R}_+^{n_1+n_2+m+m'} \mid \mu X^2 \geq x^2 \wedge \mu Z \leq z \text{ for some } \mu \in \mathbf{R}_+^p\}. \tag{4.7}$$

The first inequality in (4.7) reflects costly disposability of inputs that cause pollution and the second reflects costly disposability of pollution. Since $T_2$ is independent of $x^1$ and $y$, no inequalities need to be specified for $x^1$ and $y$.

A dataset coming from pollution-generating units must simultaneously belong to both $T_1$ and $T_2$. The overall technology that exhibits by-production is the intersection of $T_1$ and $T_2$

$$T_{BP} = \{\langle x^1,x^2,y,z\rangle \in \mathbf{R}_+^{n_1+n_2+m+m'} \mid \lambda[X^1 X^2] \leq \langle x^1,x^2\rangle, \lambda Y \geq y,$$
$$\mu X^2 \geq x^2,\ \mu Z \leq z, \text{ for some } \langle \lambda,\mu\rangle \in \mathbf{R}_+^{2p}\}. \tag{4.8}$$

The above construction of $T_{BP}$ using activity analysis involves two sets of production relations. These are reflected in the two different intensity vectors $\lambda$ and $\mu$, each of which is applied to the same dataset.

These sets for Example 1 are depicted in the first three panels of Fig. 1. Noting that $T_1$ is independent of $z$ and $T_2$ is independent of $y$, Panels 1 and 3 of Fig. 1 show the DEA constructions of projections of $T_1$ (in the space of the input and the intended output) and $T_2$ (in the space of the input and the unintended output), respectively.[24]

Panel 2 shows the combinations of intended and unintended outputs that are feasible with $x=1$, under the by-production (BP) approaches. It is clear from Panel 2 that, in the case of BP, the output possibility set has only one efficient point, $e=\langle 1,2\rangle$ (the efficient frontier of the output possibility set is a singleton). This gives the minimal level of the unintended output and the maximal level of the intended output that can be produced when $x=1$ and corresponds to efficient points $\langle 1,2\rangle$ and $\langle 1,1\rangle$ of $T_1$ and $T_2$, respectively, as seen in Panels 1 and 3.[25] On the other hand, Panel 4 shows that the efficient frontier of the output possibility set satisfying weak disposability ($obc$) has a far greater number of points. This example suggests that the efficient frontier of the output possibility set under the BP approach spans a lower dimensional space.

[24] In an abuse of notation, we also call these projections $T_1$ and $T_2$ in Fig. 1. Panels 1 and 3 of this figure are drawn under the maintained assumption of constant returns to scale.

[25] Note that, while $e$ is not a point in our artificial dataset, the data are used to infer that $e$ belongs to this set. The other points on the frontier of the output possibility set in Panel 2 reflect the assumptions of standard output free disposability in the direction of the intended output and costly disposability in the direction of pollution. (Again, an upper bound on $z$ also exists, but this bound is of no interest and is not determined by the DEA method of constructing the by-production technology.)

Consider now DMU 2 in Example 1. Its production bundle has coordinates $\langle 1,3/2\rangle$ in Panel 1 and $\langle 1,1\rangle$ in Panel 3. Any increase in $x$, holding $y$ and $z$ constant, moves the production bundle outside the technology $T_2$ and hence outside $T$, demonstrating the failure of the reduced form technology $T$ to satisfy input disposability (for a pollution-generating input).

## 5. Measuring technical efficiency

The DEA model of by-production discussed in the previous section can lend itself to many important applications. One application to a frequently studied issue in the literature is the measurement of technical efficiencies of DMUs. In this paper, we choose this application to illustrate how the by-production approach can be employed in applied work. Two conventional efficiency indexes have been extensively employed in the environmental economics literature: the output-oriented hyperbolic (HYP) index employed in the original DEA pollution study of Färe et al. [15] and the output-oriented directional-distance-function (DDF) index employed in more recent studies (*e.g.*, [14]). These indexes are "output-oriented" because they measure efficiency in (intended and unintended) output space (*i.e.*, in the output direction).

For each technology ($T=\tilde{T}_{WD},T_{BP}$) and for each decision making unit ($d=1,\ldots,p$), the output-oriented HYP efficiency index is defined by[26]

$$E_H(x^d,y^d,z^d,T)=\min_{\beta>0}\{\beta \mid \langle x^d,y^d/\beta,\beta z^d\rangle \in T\}, \tag{5.1}$$

and the output-oriented DDF index of *inefficiency* is defined by[27]

$$I_{DD}(x^d,y^d,z^d,T)=\max\{\beta \mid \langle x^d,y^d+\beta g_y,z^d-\beta g_z\rangle \in T\}, \tag{5.2}$$

where $g=\langle g_y,g_z\rangle \in \mathbf{R}_+^{m+m\prime}$ is the arbitrary (output) "direction vector." $E_H$ maps into the (0,1] interval, while $E_{DD}$ maps into $\mathbf{R}_+$. For points on the frontier of $T$, $E_H(x,y,z,T)=1$ and $I_{DD}(x,y,z,T)=0$.[28] The vectors $\langle x^d,y^d/\overset{\star}{\beta},\overset{\star}{\beta}z^d\rangle$ and $\langle x^d,\ y^d+\overset{\star}{\beta}g_y,\ z^d-\overset{\star}{\beta}g_z\rangle$, where $\overset{\star}{\beta}$ is the solution value in each case, are referred to as "reference points"; they are comparison vectors for assessing the efficiency of a particular production vector.

### *5.1. Inadequacies of conventional efficiency indexes for the by-production approach: the hyperbolic and directional-distance-function indexes*

Using our proposed BP approach under the assumptions that $T_1$ is independent of $z$ and $T_2$ is independent of $y$, the HYP and DDF (in)efficiency indexes in (5.1) and (5.2) implicitly decompose total (in)efficiency ($\beta$) into (in)efficiency in intended production ($\beta_1$) and environmental (in)efficiency ($\beta_2$):

$$E_H(x,y,z,T_{BP})=\min_{\beta>0}\{\beta \mid \langle x,y/\beta,\beta z\rangle \in T_{BP}\}=\min_{\beta>0}\{\beta \mid \langle x,y/\beta,\beta z\rangle \in T_1 \quad \text{and} \quad \langle x,y/\beta,\beta z\rangle \in T_2\}=\max\{\beta_1,\beta_2\},\ \text{where}$$

$$\beta_1=\min_{\beta>0}\{\beta \mid \langle x,y/\beta,z\rangle \in T_1\}=: E_H^1(x,y,z,T_{BP}),$$

$$\beta_2=\min_{\beta>0}\{\beta \mid \langle x,y,\beta z\rangle \in T_2\}=: E_H^2(x,y,z,T_{BP}),\ \text{and} \tag{5.3}$$

$$I_{DD}(x,y,z,T_{BP})=\max_{\beta}\{\beta \mid \langle x,y+g_y\beta,z-g_z z\rangle \in T_{BP}\}=\max_{\beta}\{\beta \mid \langle x,y+g_y\beta,z-g_z\beta\rangle \in T_1 \quad \text{and} \quad \langle x,y+g_y\beta,z-g_z\beta\rangle \in T_2\},$$

$$=\min\{\beta_1,\beta_2\},\ \text{where}$$

$$\beta_1=\max_{\beta}\{\beta \mid \langle x,y+g_y\beta,z\rangle \in T_1\}=: I_{DD}^1(x,y,z,T_{BP}) \quad \text{and}$$

$$\beta_2=\max_{\beta}\{\beta \mid \langle x,y,z-g_z\beta\rangle \in T_2\}:=I_{DD}^2(x,y,z,T_{BP}). \tag{5.4}$$

If $\max\{\beta_1,\beta_2\}=\beta_1\neq\beta_2$ for the HYP output-oriented measure of efficiency, the data point is compared to a reference point that is weakly efficient in intended production but is not weakly environmentally efficient. If $\max\{\beta_1,\beta_2\}=\beta_2\neq\beta_1$, the reference point is weakly environmentally efficient but not weakly efficient in intended production. A similar logic applies in an obvious way for the DDF measure of inefficiency. Thus, the reference points with which different DMUs are compared to measure (in)efficiency may not be even weakly efficient when the BP approach is used, and we argue below that they typically are *not* fully efficient.

---

[26] Intuitively, given a data vector of inputs and intended and unintended outputs for a DMU, the inverse of this index provides the maximum scalar amount by which the technology permits the vector of intended outputs to be scaled up and the vector of unintended outputs to be scaled down, while holding all the inputs fixed.

[27] Intuitively, given a data vector of inputs and intended and unintended outputs for a DMU and a direction vector $\langle g_y,g_z\rangle \in \mathbf{R}_+^{m+m\prime}$ of intended and unintended outputs, this index provides the maximum scalar amount by which the technology permits the vector of intended (respectively, unintended) outputs to be increased (respectively, decreased) in the direction $g_y$ (respectively, $g_z$), while holding all the inputs fixed.

[28] Note that an HYP output-oriented index of *inefficiency* can be defined by $1/E_H(x,y,z,T)$, which lies in the interval $[1,\infty)$.

Consider the quantity vector of DMU 3 in Example 1, represented by point $a = \langle a_z, a_y \rangle = \langle 2, 2/3 \rangle$ in the output possibility set corresponding to $x=1$ in Panel 2 of Fig. 1. This corresponds to points $\langle 1, 2/3 \rangle$ and $\langle 1, 2 \rangle$ in Panels 1 and 3, respectively. If the BP approach is used to measure HYP efficiency, (5.3) and Panels 1–3 show that $\beta_1 = 1/3$ and $\beta_2 = 1/2$ so that $\max\{\beta_1, \beta_2\} = \beta_2$.[29] This implies that the reference point that is being used to measure efficiency of $\langle 2, 2/3 \rangle$ is $e' = \langle 1, 4/3 \rangle$. In contrast to the fully efficient point $e$, $e'$ is environmentally efficient but not efficient in intended production. On the other hand, the HYP efficiency of $a$ using the WD approach in Panel 4 is 0.47, and the reference point is $e''$, which is technologically efficient with respect to the WD technology.[30]

Suppose that, as is common in the literature, we adopt a direction vector $g = \langle g_z, g_y \rangle = \langle 1, 1 \rangle =: \mathbf{1}$ to compute the DDF index of inefficiency for DMU 3. If the BP approach is employed, then $\beta_1$ is implicitly defined by $\frac{2}{3} + \beta_1 = 2$, so that $\beta_1 = 4/3$. Similarly, $\beta_2$ is implicitly defined by $2 - \beta_2 = 1$ so that $\beta_2 = 1$. Thus, the DDF inefficiency score of DMU 3 is $\max\{\beta_1, \beta_2\} = \beta_2 = 1$, and this leads to a reference point $\langle 1, 5/3 \rangle$ that is environmentally efficient but not efficient in intended production.

Now consider the quantity vector of DMU 2 represented by point $b = \langle 1, 3/2 \rangle$ in the output possibility set corresponding to $x=1$ in Panel 2. This point corresponds to points $\langle 1, 3/2 \rangle$ and $\langle 1, 1 \rangle$ in Panels 1 and 3, respectively. For the HYP measure, (5.3) and Panels 1–3 of Fig. 1 imply that $\beta_2 = 1$ while $\beta_1 = 3/4 < 1$. Thus, the conventional HYP measure computed using the BP approach gives DMU 2 an efficiency score $\beta = 1$ even though DMU 2 is not efficient in both the environmental and the intended output dimensions: it is only environmentally efficient.[31]

These examples illustrate a fundamental problem with the conventional measures of efficiency when using the BP approach for constructing the technology: the efficiency score for a firm may take the value 1 for HYP measures or 0 for the DDF measure even though the firm is not weakly efficient in *both* environmental and intended-output directions. In addition, the reference point, itself, with which the firm is compared may not be weakly efficient in both these dimensions, resulting in an understatement (overstatement) of overall inefficiency (efficiency).

The DDF is particularly unsuitable for use as an inefficiency index for a BP technology. It is well known that the inefficiency scores obtained from the DDF measure can be very sensitive to the choice of the direction vector $g$.[32] This sensitivity seems to be more salient in the BP approach, however, since the choice of $g$ in this context is typically tantamount to predetermining a choice between the selection of the environmental or the intended production inefficiency components as the measure of overall inefficiency.[33]

Many (in)efficiency indexes have been proposed in the literature.[34] In empirical work on pollution-generating technologies, however, HYP and DDF are among the more widely used of these conventional indexes. Given the above problems with these two indexes under the BP approach, we propose, in the next subsection, a modification of another conventional efficiency index that is better behaved for use in measuring efficiency on BP production technologies.

### 5.2. A proposed efficiency index for by-production technologies: modification of the Färe–Grosskopf–Lovell index

The previous subsection shows that the principal problem with the widely used hyperbolic and directional-distance-function efficiency indexes applied to BP technologies is the endemic understatement of the degree of inefficiency.

The index we propose for measuring efficiency on by-production technologies is motivated by the input-oriented index proposed by Färe and Lovell [17] and extended to the full $\langle$input, output$\rangle$ space for standard technologies (with no unintended outputs) by Färe et al. [10, pp. 153–154]. The key feature of this index is that the reference points it uses to assign efficiency scores to the DMUs are all efficient, in contrast to the HYP and DDF indexes, for which the reference points are all weakly efficient. In particular, this measure deems a DMU to be efficient if and only if it is *both* environmentally efficient and efficient in intended production.[35]

As our modification is minor, we continue to refer to it as the (output oriented) Färe–Grosskopf–Lovell (FGL) index and define it as follows:

$$E_{FGL}(x,y,z,T) := \frac{1}{2} \min_{\theta,\gamma} \left\{ \frac{\sum_j \theta_j}{m} + \frac{\sum_k \gamma_k}{m'} \middle| \langle x, y \oslash \theta, \gamma \otimes z \rangle \in T \right\}, \tag{5.5}$$

[29] The intuition as to why the efficiency measure chooses $\beta = \beta_2 = \max\{\beta_1, \beta_2\}$ as a full measure of output efficiency is that, while $\langle 2\beta_2, 2/3\beta_2 \rangle$ is feasible both with respect to $T_1$ and $T_2$ with $x=1$, $\langle 2\beta_1, 2/3\beta_1 \rangle$ is feasible only with respect to $T_1$ and not $T_2$, as it implies a reduction in the level of the unintended output $z$ below the minimum that $x=1$ can produce.

[30] See Appendix posted on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/ for calculation details.

[31] Similarly, it is easy to verify that the conventional DDF measure of inefficiency also gives DMU 2 an inefficiency score of 0.

[32] See, *e.g.*, Vardanyan and Noh [37] and Färe et al. [16].

[33] See Appendix posted on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/, for a further discussion of this problem in the context of Example 1.

[34] See Russell and Schworm [34] for an analysis of these indexes and their properties.

[35] This feature is attributable to the fact that the Färe–Grosskopf–Lovell index involves a maximal contraction/expansion of all inputs/outputs in coordinate-wise directions (rather than in a maximal radial or hyperbolic direction). Hence, all the slack in inputs and outputs is removed. (Of course, our output-oriented version of this index takes up all slack only in the output space, leaving the possibility of residual slack in inputs.)

where

$$y \oslash \theta = \langle y_1/\theta_1, \ldots, y_m/\theta_m \rangle \quad \text{and} \quad \gamma \otimes z = \langle \gamma_1 z_1, \ldots, \gamma_{m'} z_{m'} \rangle. \tag{5.6}$$

This index maps into the (0,1] interval and is equal to 1 if and only if the output vectors are technically efficient. In the case of BP technologies the index decomposes as follows:

$$\begin{aligned} E_{FGL}(x,y,z,T_{BP}) &:= \frac{1}{2}\min_{\theta,\gamma}\left\{\frac{\sum_j \theta_j}{m} + \frac{\sum_k \gamma_k}{m'}\middle|\langle x,y\oslash\theta,\gamma\otimes z\rangle \in T_{BP}\right\},\\ &= \frac{1}{2}\min_{\theta,\gamma}\left\{\frac{\sum_j \theta_j}{m} + \frac{\sum_k \gamma_k}{m'}\middle|\langle x,y\oslash\theta,\gamma\otimes z\rangle \in T_1 \wedge \langle x,y\oslash\theta,\gamma\otimes z\rangle \in T_2\right\},\\ &= \frac{1}{2}\min_{\theta}\left\{\frac{\sum_j \theta_j}{m}\middle|\langle x,y\oslash\theta,z\rangle \in T_1\right\} + \frac{1}{2}\min_{\gamma}\left\{\frac{\sum_k \gamma_k}{m'}\middle|\langle x,y,\gamma\otimes z\rangle \in T_2\right\},\\ &=: \frac{1}{2}[E^1_{FGL}(x,y,z,T_1)+E^2_{FGL}(x,y,z,T_2)] =: \frac{1}{2}[\beta_1+\beta_2] = \beta, \end{aligned} \tag{5.7}$$

where the third identity follows from independence of $T_1$ from $z$ and independence of $T_2$ from $y$. This index is one-half of the sum of the average maximal coordinate-wise expansions of intended-output quantities and the average maximal coordinate wise contractions of unintended-output quantities subject to the constraint that the expanded/contracted output-quantity vector remain in the production possibility set for a given input vector. Under our independence assumptions, the index decomposes into the sum of a standard intended-output-oriented index defined on $T_1$ ($\beta_1$) and an environmental index defined on $T_2$ ($\beta_2$).

The properties of this proposed index can be illustrated using the artificial data in Example 1 above. Consider first the case of DMU 3, represented by point $a$ in Panel 2 of Fig. 1, which corresponds to points $\langle 1,2/3\rangle$ and $\langle 1,2\rangle$ in Panels 1 and 3, respectively. It is clear that $E^1_{FGL}(1,2/3,2,T_1)=1/3$ and $E^2_{FGL}(1,2/3,2,T_2)=1/2$, so that $E_{FGL}(1,2/3,2,T_{BP})=5/12 < E_H(1,2/3,2,T_{BP})=1/2$. Moreover, the reference point for $a$ is the fully efficient point $e$ in Panel 2. Consider now the quantity vector of DMU 2 represented by point $b=\langle 1,3/2\rangle$ in Panel 2, which corresponds to points $\langle 1,3/2\rangle$ and $\langle 1,1\rangle$ in Panels 1 and 3, respectively. Although this point is not fully efficient, the values of both HYP and DDF are equal to 1. On the other hand, for this DMU, $E^2_{FGL}(1,3/2,1,T_2)=1$ but $E^1_{FGL}(1,3/2,1,T_1)=(3/2)/2=3/4$, so that $E_{FGL}(1,3/2,1,T_{BP})=7/8$. These examples illustrate the fact that the proposed index corrects the principal problem with the HYP and DDF indexes in the measurement of efficiency on BP technologies: a DMU is judged efficient by the FGL index if and only if it is efficient in both the environmental and intended-output directions. In particular, the FGL efficiency scores will typically be lower than the HYP efficiency scores.

It can also be verified that, for DMU 3, $E_{FGL}(1,2/3,2,T_{WD})=0.47$, and the associated reference point is $e''$ in Panel 4 of Fig. 1. Hence, the FGL efficiency score for DMU 3 under the WD approach is higher than under the BP approach. Further, $e''$ is technologically infeasible under the BP approach, while the analogous reference point $e$ for DMU 3 under the BP approach is technologically infeasible under the WD approach. The output quantity vector associated with DMU 2 is efficient under the WD approach ($E_{FGL}(1,3/2,1,T_{WD})=1$ and it involves no slack viz-a-viz the WD technology). But, as this vector is inefficient under the BP approach, it has an FGL efficiency score below 1 under this approach. Thus, FGL efficiency scores and associated reference points are typically quite different across the BP and WD approaches. In particular, if a DMU is judged efficient by the FGL index under the BP approach, it will also be judged efficient under the WD approach. But the converse is not true. Thus, the FGL efficiency scores under the WD approach will typically be at least as high as those under the BP approach.

## 6. An empirical application

We now illustrate the implementation of the efficiency indexes discussed above on a BP technology constructed with an actual database. The primary objective of this empirical exercise is to compare efficiency indexes both within BP and WD technologies and across BP and WD technologies. The comparative arguments made in Sections 4 and 5 are confirmed by this empirical exercise.

### 6.1. A brief description of the dataset

We use annual data for 92 coal-fired electric power plants from 1985 to 1995.[36] It includes only those plants for which coal constituted at least 95% of the total fuel consumption.[37] This database includes observations for one intended output: net electricity generation (in kWh); two unintended outputs: sulfur dioxide ($SO_2$) and nitrogen oxide ($NO_x$) (in short-tons); two non-polluting inputs: the capital stock and the number of employees; and three pollution-generating inputs: the heat content (in Btu) of coal, oil, and natural gas consumed at each power plant. Thus $p=92$, $m=1$, $m'=2$, $n_1=2$, and $n_2=3$. Data on the capital stock and the number of employees are derived from the U.S. Federal Energy Regulation Commission's

[36] A detailed description of the data can be found in Pasurka [29].

[37] This censoring of the data is necessitated by the DEA requirement that technologies of DMUs be homogeneous.

**Table 1**
HYP and DDF (in)efficiency indexes for BP technology.

| HYP | | | DDF | | |
|---|---|---|---|---|---|
| $\beta_1$ | $\beta_2$ | $\max\{\beta_1,\beta_2\}$ | $\beta_1 \times 10^{-4}$ | $\beta_2 \times 10^{-4}$ | $\min\{\beta_1,\beta_2\}$ |
| 1 | 0.34 | $\beta_1$ | 0 | 4.40 | $\beta_1$ |
| 1 | 0.31 | $\beta_1$ | 0 | 4.62 | $\beta_1$ |
| 0.90 | 0.42 | $\beta_1$ | 37 763 | 1.23 | $\beta_2$ |
| 0.84 | 1 | $\beta_2$ | 25 620 | 0 | $\beta_2$ |
| 0.84 | 0.69 | $\beta_1$ | 122 609 | 0.73 | $\beta_2$ |
| 0.77 | 0.45 | $\beta_1$ | 30 490 | 0.32 | $\beta_2$ |
| 0.82 | 0.83 | $\beta_2$ | 8991 | 0.03 | $\beta_2$ |
| 0.89 | 0.56 | $\beta_1$ | 12 707 | 0.19 | $\beta_2$ |
| 0.88 | 0.93 | $\beta_2$ | 51 132 | 0.08 | $\beta_2$ |
| 0.74 | 0.52 | $\beta_1$ | 12 587 | 0.09 | $\beta_2$ |

Notes: Results in this table pertain to a sample of 10 DMUs for the year 1985. The direction vector employed for computing DDF is $g=\mathbf{1}$.

Form 1 survey. Information about fuel consumption and net energy generation was obtained from the U.S. Department of Energy's (DOE) Form EIA-747 survey. This information is used by DOE to derive its estimates of $SO_2$ and $NO_x$ emissions. Emission data are provided by the U.S. Environmental Protection Agency.

### 6.2. Some points to note regarding the methodology

The various efficiency indexes are calculated by executing mathematical programming problems. In particular, the appropriate objective function in (5.1), (5.2), or (5.5) is optimized subject to the constraints in (4.4), (4.5) or (5.5), respectively; *e.g.*, to calculate the two FGL indexes, $E^1_{FGL}(x,y,z,T_1)$ and $E^2_{FGL}(x,y,z,T_2)$, respectively, on a BP technology, solve

$$\min_{\beta,\lambda} \beta \quad \text{s.t.} \quad \sum_{d=1}^{92} \lambda^d x_i^d \le x_i^{d\prime}, \quad i=1,\ldots,5 \wedge \quad \sum_{d=1}^{92} \lambda^d y^d \ge y^{d'}/\beta \quad \wedge \quad \lambda^d \ge 0,\ d=1,\ldots,92, \tag{6.1}$$

$$\min_{\gamma,\mu} \frac{\gamma_1+\gamma_2}{2} \quad \text{s.t.} \quad \sum_{d=1}^{92} \mu^d x_i^d \ge x_i^{d\prime}, \quad i=3,4,5, \wedge \quad \sum_{d=1}^{92} \mu^d z_k^d \le \gamma_k z_k^{d\prime},\ k=1,2 \quad \wedge\ \mu^d \ge 0,\ d=1,\ldots,92. \tag{6.2}$$

$E_{FGL}(x,y,z,T_{BP})$ is then obtained as the simple average of the two value functions for these optimization problems.[38]

### 6.3. Results

Table 1 reports the (in)efficiency scores of a randomly chosen sample of 10 DMUs for the year 1985 under the BP approach. The results depicted in Table 1 underscore the sensitivity of the DDF measure to the choice of the direction vector (illustrated above using Example 1). In our dataset, the consequence of choosing $g=\langle 1,1,1\rangle =: \mathbf{1}$ (which is a popular choice in the literature) is that the DDF measure of inefficiency picks up the environmental inefficiency component as the overall measure for most DMUs. The magnitudes of the HYP efficiency figures for $\beta_1$ and $\beta_2$ for these firms are reasonably comparable (ranging from 0.7416 to 1.000 for $\beta_1$ and from 0.3052 to 1.000 for $\beta_2$), so that the operation $\beta=\max\{\beta_1,\beta_2\}$ is, in some sense, non-discriminatory in choosing between $\beta_1$ and $\beta_2$. The magnitudes of $\beta_1$ and $\beta_2$ for the DDF measure, however, are in orders ranging from $10^8$ to $10^{10}$ and from $10^3$ to $10^5$, respectively, so that, except when $\beta_1=0$, the operation $\beta=\min\{\beta_1,\beta_2\}$ predominantly favors $\beta_2$ over $\beta_1$. Primarily for this reason, we do not present further results for the DDF measure of inefficiency.

Table 2 contains the mean values of the HYP and FGL efficiency indexes for each year in our sample. Columns (1) and (2) pertain to the WD technology and Columns (3)–(8) pertain to the BP technology. The BP approach is our proposed method of constructing pollution-generating technologies and the FGL index is our proposed method of calculating efficiency on BP technologies.

Columns (1) and (2) and Columns (5) and (8) of Table 2 show that, under both the WD and BP approaches, the HYP index runs higher than the FGL index. As in Example 1, this comparison reflects the fact that the expansion/contraction to the frontier of the latter takes up all the slack in outputs, thus comparing the output quantity vector to a reference vector on the efficient frontier, whereas the expansion/contraction of the former leaves some slack, comparing the output quantity vector to a point on the frontier but not necessarily in its efficient subset.

[38] With a single intended output, these programs are conveniently linear. The computer code for all calculations on our dataset can be found on the Journal's online repository of supplementary material, accessible at http://www.aere.org/journals/.

**Table 2**
Mean efficiency values.

| Year | WD technology | | BP technology | | | | | |
|---|---|---|---|---|---|---|---|---|
| | HYP | FGL | HYP | | | FGL | | |
| | (1) $\beta$ | (2) $\beta$ | (3) $\beta_1$ | (4) $\beta_2$ | (5) $\beta$ | (6) $\beta_1$ | (7) $\beta_2$ | (8) $\beta$ |
| 1985 | 0.94 | 0.78 | 0.89 | 0.64 | 0.90 | 0.89 | 0.52 | 0.70 |
| 1986 | 0.94 | 0.78 | 0.87 | 0.62 | 0.88 | 0.87 | 0.49 | 0.68 |
| 1987 | 0.95 | 0.79 | 0.90 | 0.65 | 0.92 | 0.90 | 0.54 | 0.72 |
| 1988 | 0.95 | 0.81 | 0.88 | 0.63 | 0.90 | 0.88 | 0.60 | 0.74 |
| 1989 | 0.95 | 0.82 | 0.90 | 0.63 | 0.92 | 0.90 | 0.60 | 0.75 |
| 1990 | 0.94 | 0.82 | 0.88 | 0.62 | 0.91 | 0.88 | 0.59 | 0.74 |
| 1991 | 0.95 | 0.80 | 0.89 | 0.59 | 0.91 | 0.89 | 0.54 | 0.71 |
| 1992 | 0.95 | 0.79 | 0.89 | 0.58 | 0.91 | 0.89 | 0.53 | 0.71 |
| 1993 | 0.95 | 0.79 | 0.89 | 0.60 | 0.91 | 0.89 | 0.54 | 0.72 |
| 1994 | 0.94 | 0.77 | 0.88 | 0.60 | 0.90 | 0.88 | 0.56 | 0.72 |
| 1995 | 0.91 | 0.74 | 0.80 | 0.61 | 0.84 | 0.80 | 0.55 | 0.68 |

**Table 3**
Counts of (weakly) efficient DMUs.

| Year | WD technology | | BP technology | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | HYP | FGL | HYP | | | | FGL | | | |
| | (1) $\beta=1$ | (2) $\beta=1$ | (3) $\beta_1=1$ | (4) $\beta_2=1$ | (5) $\beta_1=1,\ \beta_2=1$ | (6) $\beta=1$ | (7) $\beta_1=1$ | (8) $\beta_2=1$ | (9) $\beta_1=1,\ \beta_2=1$ | (10) $\beta=1$ |
| 1985 | 35 | 3 | 9 | 9 | 1 | 17 | 9 | 4 | 0 | 0 |
| 1986 | 36 | 3 | 5 | 6 | 1 | 10 | 5 | 4 | 0 | 0 |
| 1987 | 43 | 3 | 12 | 10 | 1 | 11 | 12 | 6 | 0 | 0 |
| 1988 | 41 | 7 | 8 | 8 | 0 | 16 | 8 | 5 | 0 | 0 |
| 1989 | 41 | 6 | 9 | 11 | 1 | 19 | 9 | 9 | 0 | 0 |
| 1990 | 36 | 6 | 7 | 11 | 0 | 18 | 7 | 8 | 0 | 0 |
| 1991 | 39 | 4 | 8 | 10 | 2 | 16 | 8 | 7 | 1 | 1 |
| 1992 | 38 | 5 | 10 | 8 | 1 | 17 | 10 | 7 | 1 | 1 |
| 1993 | 44 | 5 | 7 | 7 | 0 | 14 | 7 | 5 | 0 | 0 |
| 1994 | 43 | 3 | 6 | 6 | 0 | 12 | 6 | 5 | 0 | 0 |
| 1995 | 34 | 3 | 9 | 9 | 0 | 18 | 9 | 5 | 0 | 0 |

Table 2 also indicates that, for our dataset, both the HYP and FGL efficiency estimates are consistently higher for the WD technology than for the BP technology, a phenomenon that we explained above using Example 1. These differences in the efficiency scores across the BP and WD technologies suggest that, for both HYP and FGL measures, the reference points with respect to which efficiency is measured are different under the two approaches. In particular, in the FGL case, all the reference points are efficient, whereas for the HYP case, all are only weakly efficient. Thus, our results show that the sets of efficient and the sets of weakly efficient points differ across WD and BP technologies.

In the case of our particular dataset, regardless of the index used, Table 2 also shows that the degree of inefficiency in the pollution technology $T_2$ is much larger than that in the intended-production technology $T_1$: apparently, the DMUs in our dataset are less concerned about the environmental dimension of their production activities or environmental efficiency is more difficult to achieve.

The FGL index records greater pollution-generation inefficiency than does the HYP index. An obvious explanation could again be the differences in the way in which the two indexes treat slacks in outputs.[39]

Table 3 provides counts of weakly efficient and efficient firms using the HYP and FGL indexes, respectively, for the two technologies. Columns (1) and (6) and Columns (2) and (10) provide a comparison across WD and BP technological specifications of numbers of firms that receive an efficiency score of 1 under the HYP and FGL measures, respectively. The table shows that, for both the HYP and FGL indexes, the WD technological specification results in a larger number of firms receiving an efficiency score 1 than does the BP technological specification. This seems consistent with the findings from Example 1: the frontier of the output possibility set is larger under the WD specification than under the BP specification.

[39] Note that the (output oriented) HYP and FGL indexes take the same values for the intended-production technology $T_1$ because, with only a single intended output, they collapse to the same index.

**Table 4**
Spearman rank correlation coefficients among efficiency indexes.

| Year | Across BP and WD technologies | | Within BP technology | | | | |
|---|---|---|---|---|---|---|---|
| | | | $\rho$(HYP,FGL) | | | $\rho(\beta_1,\beta_2)$ | |
| | (1) HYP | (2) FGL | (3) $\beta$ | (4) $\beta_1$ | (5) $\beta_2$ | (6) HYP | (7) FGL |
| 1985 | 0.71 | 0.82 | 0.60 | 1.00 | 0.89 | −0.08 | −0.01 |
| 1986 | 0.70 | 0.89 | 0.53 | 1.00 | 0.87 | −0.12 | −0.09 |
| 1987 | 0.60 | 0.78 | 0.54 | 1.00 | 0.91 | −0.13 | −0.12 |
| 1988 | 0.60 | 0.77 | 0.42 | 1.00 | 0.97 | −0.23 | −0.23 |
| 1989 | 0.63 | 0.66 | 0.45 | 1.00 | 0.99 | −0.28 | −0.27 |
| 1990 | 0.58 | 0.71 | 0.50 | 1.00 | 0.98 | −0.24 | −0.24 |
| 1991 | 0.52 | 0.79 | 0.46 | 1.00 | 0.96 | −0.20 | −0.17 |
| 1992 | 0.57 | 0.87 | 0.43 | 1.00 | 0.94 | −0.21 | −0.13 |
| 1993 | 0.50 | 0.82 | 0.42 | 1.00 | 0.94 | −0.18 | −0.18 |
| 1994 | 0.54 | 0.76 | 0.47 | 1.00 | 0.96 | −0.13 | −0.16 |
| 1995 | 0.59 | 0.78 | 0.72 | 1.00 | 0.96 | −0.18 | −0.14 |

Hence, the probability of a DMU being assigned an efficiency value of 1 is greater under the WD approach than under the BP approach.

Columns (3)–(10) of Table 3 help to compare the performance of FGL and HYP indexes under the BP approach. First, it is not surprising that the HYP index, which allows slack to remain in reference output vectors, judges at least as many DMUs to be efficient (environmentally, in intended production, and overall) as does the FGL measure. This comparison is indicated by comparing Column (3) with Column (7), Column (4) with Column (8), and Column (6) with Column (10).[40] Second, it follows that the set of DMUs judged environmentally efficient by FGL is a subset of the DMUs judged environmentally efficient by HYP. Finally, as demonstrated by Example 1, the HYP index gives efficiency score 1 to DMUs that are efficient in intended outputs *or* are environmentally efficient *or* are both. Hence, Column (6) can also be equivalently obtained by adding Columns (3) and (4) and subtracting Column (5) from this sum. On the other hand, as also demonstrated by Example 1, FGL is more demanding in judging a DMU efficient: it gives efficiency score 1 to a DMU if and only if it is efficient both environmentally and in intended production. Thus, Column (10) is equal to Column (9).

Table 4 shows how the rankings of firms on the basis of their efficiency scores compare across the two efficiency indexes HYP and FGL, across the two technological specifications, and across the environmental and intended-production efficiency scores. Columns (1) and (2) of Table 4 show that, for both HYP and FGL, the Spearman correlation coefficients between the efficiency scores under the WD and BP approaches are moderately high and positive: the rank correlation coefficients lie in the range 0.50–0.71 and 0.66–0.89 for the HYP and FGL measures, respectively. In the light of the significant conceptual differences between the two approaches (in particular, the differences in the frontiers of the BP and WD technologies), which are reinforced strongly by our empirical findings above, the BP approach seems to make a larger difference in the levels than in the ranking of the efficiency scores of the DMUs.

Table 4 also allows comparison of rankings under the HYP and FGL indexes applied to BP technologies. Given that in our dataset there is only a single intended output, there are no differences in the efficiency scores for intended production obtained from the HYP and FGL measures. Hence, the Spearman correlation coefficients in Column (4) are all equal to 1. Our dataset also exhibits high rank correlations between environmental efficiency scores obtained from the FGL and HYP measures: as seen in Column (5), the rank correlation coefficients lie in the range 0.87–0.99. Nevertheless, the rank correlation coefficients between overall efficiency scores obtained under FGL and HYP are on the lower side: as seen in Column 3, these lie in the range 0.42–0.72. This could be explained by the differences in the way HYP and FGL indexes aggregate over environmental and intended output efficiency scores. In Example 1, we saw that the HYP gives an efficiency score of 1 to a DMU that is environmentally efficient but not efficient in intended production or vice versa. The FLG index, however, penalizes such DMUs for the slack in production of the intended or the unintended output and gives them a lower score.

Columns (6) and (7) of Table 4 show the rank correlation coefficients between efficiency scores in intended and unintended productions for the HYP and FGL indexes under the BP approach. These values are all negative and low; the Spearman correlation coefficients range between −0.08 to −0.28 and −0.01 to −0.27 for the HYP and FGL indexes, respectively. Negative correlation values indicate that DMUs are more efficient in intended production are likely to be more environmentally inefficient, and vice versa. This may suggest that the DMUs face some trade-offs between efficiency in intended production and in pollution generation. In our dataset, however, these trade-offs are weak, as the correlation values are very low. Thus, one may conclude that most DMUs in our dataset do not face significant trade-offs between intended production and residual generation and can improve simultaneously on both environmental and intended-output efficiencies.

[40] With respect to the intended-production technology $T_1$, since there is only one intended output, there is no slack remaining in the reference vector when the HYP index gives a DMU an efficiency score of 1. Hence, Columns (3) and (7), are identical.

## 7. By-production versus weak disposability: comparisons of DEA formulations in the presence of abatement efforts

The WD approach explains the positive correlation between intended outputs and pollution through abatement efforts of firms that are not modeled. Hence, it considers only a reduced form of the overall technology in the space of inputs and all unintended and intended outputs other than the abatement output. In this section, we extend the DEA formulation of a BP technology to include abatement efforts made by firms and derive the DEA analog of its reduced form defined in (3.11). With the help of an example, we then compare the reduced forms of the two technologies.

A DEA version of the BP technology in the presence of an abatement output is derived as follows. With respect to the intended technology $T_1$, abatement is a standard output that satisfies standard output free disposability. The residual-generating mechanism $T_2$, on the other hand, satisfies costly disposability of abatement output. Thus,

$T_{BP} = T_1 \cap T_2$, where

$$T_1 = \{\langle x^1, x^2, y, y^a, z\rangle \in \mathbf{R}_+^{n_1+n_2+m+1+m'} \mid \lambda[X^1\ X^2] \le \langle x^1,\ x^2\rangle \wedge \lambda Y \ge y \wedge \lambda A \ge y^a \text{ for some } \lambda \in \mathbf{R}_+^p\}, \text{ and}$$

$$T_2 = \{\langle x^1, x^2, y, y^a, z\rangle \in \mathbf{R}_+^{n_1+n_2+m+1+m'} \mid \mu X^2 \ge x^2 \wedge \mu A \le y^a \wedge \mu Z \le z \text{ for some } \mu \in \mathbf{R}_+^p\}, \tag{7.1}$$

where $A$ is the vector of abatement outputs for the $p$ firms.

Holding all input quantities fixed at $x$, we next derive a DEA version of the reduced form of $T_{BP}$. Precisely, this is the projection of the output possibility set of $T_{BP}$ (corresponding to input-quantity level $x$) defined in the $\langle z, y, y^a\rangle$ space into the $\langle z, y\rangle$ space.

Noting that technology $T_1$ is independent of $z$, the DEA construction of the projection of the output-possibility set for technology $T_1$ (corresponding to input level $x$) into the $\langle y^a, y\rangle$ space is denoted by $\hat{P}_1(x)$. In a similar manner, noting that technology $T_2$ is independent of $y$, we define the DEA construction of the projection $\hat{P}_2(x)$ of $T_2$ into the $\langle y^a, z\rangle$ space.

The DEA versions of the reduced form of $T_{BP}$ and the WD technology (see (4.4) in the $\langle z, y\rangle$ space, for a fixed level $x$ of input quantities, are defined as follows:

$$\hat{P}_{BP}(x) = \{\langle z, y\rangle \in \mathbf{R}_+^{m+m\prime} \mid \exists\, y^a \in \mathbf{R}_+ \text{ such that } \langle y^a, y\rangle \in \hat{P}_1(x) \wedge \langle y^a, z\rangle \in \hat{P}_2(x)\},$$

$$\tilde{P}_{WD}(x) = \{\langle z, y\rangle \in \mathbf{R}_+^{m+m\prime} \mid \langle x, y, z\rangle \in \tilde{T}_{WD}\}. \tag{7.2}$$

In Example 41 below, we compare $\hat{P}_{BP}(x)$ and $\tilde{P}_{WD}(x)$.

**Example 2.** $p=8$, $n_2=1$, $n_1=0$, $m=m'=1$, and $x=1$. The (artificial) dataset is

| DMU | $x$ | $y^a$ | $y$ | $z$ | DMU | $x$ | $y^a$ | $y$ | $z$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 8 | 9 | 5 | 1 | 4 | 1 | 2 |
| 2 | 1 | 1 | 7 | 6 | 6 | 1 | 5 | 4 | 0 |
| 3 | 1 | 2 | 6 | 8 | 7 | 1 | 6 | 2 | 0 |
| 4 | 1 | 3 | 6 | 3 | 8 | 1 | 7 | 1 | 11 |

(7.3)

After plotting the data, we find that $\hat{P}_1(1)$ and $\hat{P}_2(1)$ can be represented functionally by piece-wise linear functions

$$\begin{aligned}\psi^1(y^a) &= 8-\tfrac{2}{3}y^a \text{ if } y^a \in [0,3] & \psi^2(y^a) &= 9-3y^a \text{ if } y^a \in [0,1].\\ &= 9-y^a \text{ if } y^a \in [3,5] & &= \tfrac{15}{2}-\tfrac{3}{2}y^a \text{ if } y^a \in [1,5],\\ &= \tfrac{23}{2}-\tfrac{3}{2}y^a \text{ if } y^a \in [5,7] & &= 0 \text{ if } y^a \ge 5.\end{aligned} \tag{7.4}$$

The sets $\hat{P}_1(1)$ and $\hat{P}_2(1)$ are shown in Panels 1 and 2 of Fig. 2. (7.2) implies that $\hat{P}_{BP}(1)$ (shown in Panel 3 of Fig. 2) is constructed as follows:

$$\hat{P}_{BP}(1) = \{\langle z, y\rangle \in \mathbf{R}_+^2 \mid z \ge \psi^2(y^a) \wedge y \le \psi^1(y^a) \wedge y^a \in [0,7]\}. \tag{7.5}$$

Note that the construction of $\hat{P}_{BP}(1)$ involves explicit reference to the abatement output.[41] No reference was made, however, to data on $y^a$ in the DEA construction of $\tilde{P}_{WD}(1)$ in Panel 4 of Fig. 2.

Moreover, while weak disposability holds for $\tilde{P}_{WD}(1)$, the data are such that null-jointness is violated. This violation can be rationalized by the possibility that the abatement output of a firm can completely mitigate (net) pollution even when it is producing positive (gross) amounts of the intended outputs. (Consider abatement activities like recycling of wastes or the possibility that wastes are biodegradable and hence can be completely eliminated using only "clean" inputs like labor.) Note that null-jointness is *not* imposed on the empirical DEA construction of the technology (4.4): if the dataset were to contain an observation with a positive intended output and no positive unintended output, the empirically constructed WD technology would violate null jointness, and if it were not to contain such a point, it would satisfy this assumption. The same thing can be said about our empirical construction of the BP technology. Thus, the satisfaction of null-jointness for a particular construction of a WD or BP technology is entirely an empirical issue.

[41] In particular, we have been able to express the frontier of $\hat{P}_{BP}(1)$ as a vector-valued function of $y^a$.

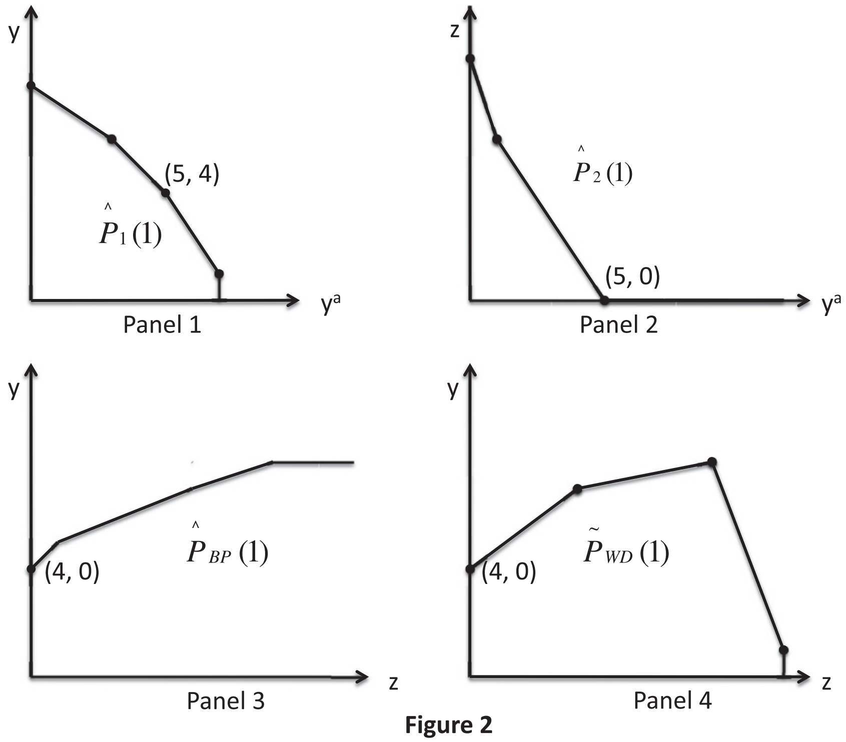


**Fig. 2.**

Finally, the boundary of $\tilde{P}_{WD}(1)$ has a negatively sloped region, indicating a negative correlation between intended and unintended outputs in that region. The frontier of $\hat{P}_{BP}(1)$, on the other hand, is everywhere non-negatively sloped.

## 8. Conclusions

Pollution is an unintended output that cannot be freely disposed of. Underlying its production is a set of chemical and physical reactions that take place in nature when firms engage in the production of intended outputs. These natural reactions define nature's residual generation mechanism, which is a relation between the residuals generated and some inputs that are used or some intended outputs that are produced by the firm: hence, the inevitability of a certain minimal amount of pollution being generated when firms engage in intended production. We call this phenomenon by-production of pollution. The larger is the scale of intended production, the greater are the pollution-causing inputs being used or the greater are the pollution-causing intended outputs being produced, and hence, the greater the generation of pollution. This provides the fundamental explanation for the positive correlation that is observed between intended production and residual generation.

Standard approaches in the existing literature, on the other hand, usually attribute the observed positive correlation between pollution generation and intended production to resource-costly abatement options of firms. Such options, however, are not explicitly modeled, and only a reduced form of the technology is considered. Pollution is either treated as an input satisfying standard input free disposability or is considered as an output that is weakly disposable.

To capture the phenomenon of by-production, we model pollution-generating technologies as a composition of two technologies: an intended-production technology and a residual-generation technology. The former describes how inputs are transformed into intended outputs, is assumed to be independent of the level of pollution, and satisfies standard free-disposability properties. The latter reflects nature's residual generation, violates standard disposability properties with respect to goods that result in (affect) pollution generation, and exhibits costly disposability with respect to pollution. As a result, the overall technology violates standard disposability with respect to inputs that cause (affect) pollution generation and exhibits costly disposability with respect to pollution. In these ways, a by-production technology, which is based on multiple production relations, is different and better able to capture the observed trade-offs in production than the usual input and output approaches to modeling pollution-generating technologies based on just a single production relation.

We formulate DEA specifications of technologies that satisfy by-production, with or without pollution-abatement activities, and employ them to measure technical efficiency of firms. In the context of by-production, standard measures of efficiency decompose very naturally into environmental and intended-output efficiencies. However, we find that, in the context of by-production, the commonly used indexes of (in)efficiency, the hyperbolic and the directional-distance-function index, overstate efficiency. In the existing set of (in)efficiency indexes proposed in the literature, we find that a modification of an index proposed by Färe et al. [10] corrects the flaws in the hyperbolic and directional-distance-function indexes for measurement of efficiency on by-production technologies. A comparison of the values of this index with those of the hyperbolic and directional-distance-function indexes, using a database for electric power firms, supports our arguments about the inadequacies of the latter.

## References

[1] R.U. Ayres, A.V. Kneese, Production, consumption and externalities, American Economic Review 59 (1969) 282–297.
[2] A.J. Barbera, V.D. McConnell, The impact of environmental regulation on industry productivity: direct and indirect effects, Journal of Environmental Economics and Management 18 (1990) 50–65.
[3] W.J. Baumol, W.E. Oates, The Theory of Environmental Policy, Second ed., Cambridge University Press Boyd, Cambridge, 1988.
[4] G.A. Boyd, J.D. McClelland, The impact of environmental constraints on productivity improvement in integrated paper plants, Journal of Environmental Economics and Management 38 (1999) 121–142.
[5] R.G. Chambers, T. Melkonyan, Production Technologies, Material Balance, and the Income-Environmental Quality Trade-Off, University of Exeter Working Paper, 2012.
[6] T. Coelli, L. Lauwers, G. Van Huylenbroeck, Environmental efficiency measurement and the materials balance condition, Journal of Productivity Analysis 28 (2007) 3–12.
[7] J.S. Coggins, J.R. Swinton, The price of pollution: a dual approach to valuing $SO_2$ allowances, Journal of Environmental Economics and Management 30 (1996) 58–72.
[8] M.L. Cropper, W.E. Oates, Environmental economics: a survey, Journal of Economic Literature 30 (1992) 675–740.
[9] R. Färe, S. Grosskopf, Measuring output efficiency, European Journal of Operational Research 13 (1983) 173–179.
[10] R. Färe, S. Grosskopf, C.A. Knox Lovell, The Measurement of Efficiency of Production, Kluwer-Nijhoff, Boston, 1985.
[11] R. Färe, S. Grosskopf, C.A. Knox Lovell, Production Frontiers, Cambridge University Press, Cambridge, 1994.
[12] R. Färe, S. Grosskopf, C.A. Knox Lovell, C. Pasurka, Multilateral productivity comparisons when some outputs are undesirable: a nonparametric approach, The Review of Economics and Statistics 71 (1989) 90–98.
[13] R. Färe, S. Grosskopf, C.A. Knox Lovell, S. Yaisawarng, Derivation of shadow prices for undesirable outputs: a distance function approach, The Review of Economics and Statistics 75 (1993) 374–380.
[14] R. Färe, S. Grosskopf, D.-W. Noh, W. Weber, Characteristics of a polluting technology: theory and practice, Journal of Econometrics 126 (2005) 469–492.
[15] R. Färe, S. Grosskopf, C. Pasurka, Effects of relative efficiency in electric power generation due to environmental controls, Resources and Energy 8 (1986) 167–184.
[16] R. Färe, S. Grosskopf, C. Pasurka, Environmental production functions and environmental directional distance functions, Energy 32 (2007) 1055–1066.
[17] R. Färe, C.A.K. Lovell, Measuring the technical efficiency of production, Journal of Economic Theory 19 (1978) 150–162.
[18] F. Førsund, Good modelling of bad outputs: pollution and multiple-output production, International Review of Environmental and Resource Economics 3 (2009) 1–38.
[19] H. Fried, C.A. Knox Lovell, S.S. Schmidt, The Measurement of Productive Efficiency and Growth, Cambridge University Press, Cambridge, 2008.
[20] R. Frisch, Theory of Production, D. Reidel Publishing Company, Dordrecht, 1965.
[21] A. Hailu, T.S. Veeman, Environmentally sensitive productive analysis of the Canadian Pulp and paper industry, 1959–1994: an input distance function approach, Journal of Environmental Economics and Management 40 (2000) 251–274.
[22] Z. Hua, Y. Bien, Performance measurement for network DEA with undesirable factors, International Journal of Management and Decision Making 9 (2008) 141–152.
[23] M.N. Murty, S. Kumar, Measuring cost of environmentally sustainable industrial development in India: a distance function approach, Environment and Development Economics 7 (2002) 467–486.
[24] M.N. Murty, S. Kumar, Win-win opportunities and environmental regulation: testing of porter hypothesis for Indian manufacturing industries, Journal of Environmental Management 67 (2003) 139–144.
[25] M.N. Murty, S. Kumar, M. Paul, Environmental regulation, productive efficiency, and cost of pollution abatement: a case study of sugar industry in India, Journal of Environmental Management 79 (2006) 1–9.
[26] S. Murty, Externalities and fundamental nonconvexities: a reconciliation of approaches to general equilibrium externality modeling and implications for decentralization, Journal of Economic Theory 145 (2010) 331–353.
[27] S. Murty, On the theory of a firm: the case of by-production of emissions, The Warwick Economics Research Papers Series (TWERPS), No. 934, 2010.
[28] S. Murty, R.R. Russell, On modeling pollution-generating technologies, Department of Economics, University of California, Riverside, Discussion Papers Series, No. 02-14, 2002.
[29] C.A. Pasurka, Decomposing electric power plant emissions within a joint production framework, Energy Economics 28 (2006) 26–43.
[30] R. Pethig, Non-linear production, abatement, pollution, and materials balance reconsidered, Journal of Environmental Management 51 (2006) 185–204.
[31] R.W. Pittman, Multilateral productivity comparisons with undesirable outputs, The Economic Journal 93 (1983) 391–883.
[32] S. Reinhard, C.A. Knox Lovell, G.J. Thijssen, Econometric estimation of technical and environmental efficiency: an application to Dutch dairy farms, American Journal of Agricultural Economics 81 (1999) 44–60.
[33] S. Reinhard, C.A. Knox Lovell, G.J. Thijssen, Environmental efficiency with multiple environmentally detrimental variables; estimated with SFA and DEA, European Journal of Operational Research 121 (2000) 287–303.
[34] R.R. Russell, W. Schworm, Properties of inefficiency indexes on $\langle$Input, Output$\rangle$ space, Journal of Productivity Analysis 36 (2011) 143–156.
[35] R.W. Shephard, Cost and Production Functions, Princeton University Press, Princeton, 1953.
[36] R.W. Shephard, Indirect Production Functions, Verlag Anton Hain, Meisenheim Am Glan, 1974.
[37] M. Vardanyan, D.-W. Noh, Approximating pollution abatement costs via alternative specifications of a multi-output production technology: a case of US electric utility industry, Journal of Environmental Management 80 (2006) 177–190.
[38] P. Zhou, B.W. Ang, K.-L. Poh, A survey of data envelopment analysis in energy and environmental studies, European Journal of Operational Research 189 (2008) 1–18.